Bayesian Hierarchical Models to Augment the Mediterranean Ocean Forecast System

ONR Contract N00014-05-C-0198 to NorthWest Research Associates

Second Progress Report, First Anniversary of Funding: 15 June 2006

Introduction

The first full year of research for the project entitled "Bayesian Hierarchical Models (BHM) to Augment the Mediterranean Ocean Forecast System (MFS)" completed at the end of May 2006. Project achievements have met or exceeded plans put forth in the proposal. Prof. Nadia Pinardi (Univ. Bologna, INGV) and Dr. Ralph F. Milliff (NWRA/CoRA) presented early results to Physical Oceanography Program Managers in a seminar at ONR Headquarters, in Arlington, in early May.[1] This annual report reviews highlights from that presentation.

MFS-Wind-BHM

Initial project activities focused on the development of a BHM for the surface wind process, based on the stochastic geostrophy model first proposed by Royle et al. (1998). The MFS-Wind-BHM implementation was led by Prof. Chris Wikle (U. Missouri, project Co-PI). Several versions of the surface wind BHM have been implemented and tested. Version 4 has been implemented in the first MFS ensemble data assimilation and forecast experiments reviewed here.

In probability model notation, MFS-Wind-BHM is given at the lowest (i.e. most general) level in two parts; a data stage distribution and a process model stage distribution. These are:

$$\prod_{t=1}^{T}[S_t|W_t,\theta_S] \times \prod_{t=1}^{T}[A_t^w|W_t,\theta_{AW}][A_t^p|P_t,\theta_{AP}] \times \prod_{t=T+1}^{T+L}[F_t^w|W_t,\theta_{FW}][F_t^p|P_t,\theta_{FP}]; \tag{1}$$

and

$$[b_0] \times \prod_{t=1}^{T+L}[W_t|P_t,\theta_W][P_t|b_t,\theta_P][b_t|b_{t-1},\theta_b] \tag{2}$$

where probability distributions are noted by square brackets, and, for example, the conditional distribution for the wind process W^t given the SLP process P^t is denoted $[W^t|P^t]$. In (1) the data stage distributions for the data assimilation period $t = 1, T$ include scatterometer winds S, and

[1] CDs containing the powerpoint slides used in the seminar have been sent to Drs. Scott Harper, Manuel Fiadeiro and Theresa Paluszkiewicz, Physical Oceanography Program, ONR.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)* 15-06-2006	**2. REPORT TYPE** Annual Report	**3. DATES COVERED** *(From - To)* 5/11/05 - 5/30/06

4. TITLE AND SUBTITLE Bayesian Hierarchical Models to Augment the Mediterranean Ocean Forecast System	**5a. CONTRACT NUMBER** N00014-05-C-0198 **5b. GRANT NUMBER** **5c. PROGRAM ELEMENT NUMBER**
6. AUTHOR(S) Dr. Ralph Milliff	**5d. PROJECT NUMBER** ONR C220P **5e. TASK NUMBER** **5f. WORK UNIT NUMBER**
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) NorthWest Research Associates 14508 NE 20th St. PO Box 3027 Bellevue, WA 98009-3027	**8. PERFORMING ORGANIZATION REPORT NUMBER** ONR C220P Annual Report #1
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) Office of Naval Research Attn: Manuel Fiadeiro, ONR 322PO Ballston Tower One 800 North Quincy Street Arlington, VA 22217-5660	**10. SPONSOR/MONITOR'S ACRONYM(S)** **11. SPONSOR/MONITOR'S REPORT NUMBER(S)**

12. DISTRIBUTION / AVAILABILITY STATEMENT

Public distribution/availability

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Project goals, as set forth in the proposal, have been met or exceeded in year 1, which ended 30 May 2006. Interactions between statisticians, oceanographers, project and MFS personnel, have resulted in robust reforecast experimental procedures, and the results to date. These results were reported in a seminar to ONR Physical Oceanography program managers. An MFSWind-BHM has been implemented in a series of ensemble reforecast experiments. Realizations from the MFS-Wind-BHM posterior distribution are used in ensemble data assimilation, and to drive members of an ensemble forecast. Results from the reforecast experiments demonstrate that initial condition uncertainty, as quantified by ensemble spread at the end of the sequential data assimilation step, is concentrated in mesoscale features of the ocean response to BHM winds. The uncertainty in the mesoscale is an appropriate property for initial conditions in an ensemble forecast system. Mesoscale uncertainty persists in the 10-day ensemble reforecasts. Differing mesoscale evolutions lead to significant differences in deep water formation for the same forecast period.

15. SUBJECT TERMS

Bayesian Hierarchical Model, Mediterranean Forecast System, data assimilation, ensemble initial conditions, ensemble forecast

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT Unclassified	**b. ABSTRACT** Unclassified	**c. THIS PAGE** Unclassified	UL	9	**19b. TELEPHONE NUMBER** *(include area code)*

analysis surface wind A^w and SLP A^p. During the forecast period $t = T, T + L$, the ECMWF forecast winds (F^w) and SLP (F^p) degrade in skill with increasing forecast lead time. Following Royle et al. (1998), the process model part of (2) is based, in part, on a *stochastic* geostrophy model, wherein the surface wind distribution $[W^t]$ is conditioned on gradients of the SLP process P^t (both cross- and down-gradient terms). SLP fields are constructed by time-dependent weightings b_t of spatial structure functions (i.e. eigenvectors). Time dependence in the prior is prescribed in an autoregressive model at the next higher level of the BHM hierarchy (not shown).

Figure 1 depicts a snapshot of 10 realizations from the posterior distribution for the surface wind process as generated in the version 4 implementation of MFS-Wind-BHM. A cluster of 10 blue surface wind vectors is depicted at every other BHM prediction grid location (prediction grid resolution is $0.25°$) in the western basin of the Mediterranean Sea. A green vector within each cluster depicts the ensemble mean surface wind for each location (i.e. the average speed and direction of the 10 realizations). A red vector (sometimes obscured) depicts the ECMWF analysis vector for the same output time. When MFS-Wind-BHM winds are organized by a prevailing weather pattern (e.g. the Sirocco event depicted in Fig. 1), the distributions of wind speeds and directions are relatively tightly spread about the ensemble mean vector. Conversely, in low wind speed events (e.g. see the Tyrrhenian Sea), the wind direction is arbitrary as reflected in the distributions that span a wide range of directions, all at very low wind speeds.

An animation has been created, in the format of Fig. 1, for every MFS-Wind-BHM output time (4-times daily) spanning the data assimilation and forecast periods (i.e. $t = 1, T + L$). During the assimilation period, there are instances when the ECMWF analysis vector departs from the MFS-Wind-BHM cluster. This is due to the influence of S in the data stage. Evidence has shown that the scatterometer winds do not influence the ECMWF analyses at the resolution of the oceanic mesoscale in the Mediterranean Sea (Milliff and Morzel, 2003). Conversely, during the forecast period, departures of the MFS-Wind-BHM clusters from the (retrospective) ECMWF analysis vectors are probably due to diminishing skill in the MFS-Wind-BHM forecast (i.e. for lead times greater than about 3 days).

The MFS-Wind-BHM animation was shown and discussed in the seminar at ONR HQ.

MFS Ensemble Initial Condition and Forecast Responses

The MFS data assimilation and forecast cycle are depicted schematically in Figure 2. Ocean data from the multi-platform MFS observing system are assimilated over a 14 day period (i.e. $T = 14d$). Data sources include XBT profiles from Volunteer Observing Ships, CTD profiles from occasional directed research cruises (including a glider experiment in the Ionian Sea), profile and current information from ARGO drifting buoys, ocean and meteorological data from fixed moorings (i.e. M3A buoys), and sea-surface height (SSH) anomaly fields from accumulated altimeter data. The ocean data are assimilated in a Reduced Order Optimal Interpolation (ROOI) procedure that uses ECMWF analysis atmospheric forcing data to force tangent linear dynamics. A 10-day ($L = 10d$) ocean forecast is launched from the end of the assimilation period. Surface forcing for the MFS forecast period is taken from ECMWF forecasts.

For thirteen days after the ocean assimilation step, an ocean simulation is run to span the period from the last ocean analysis to the present day (i.e. simulations of 1, 2, 3, ... 13 days are run on second, third, fourth, ..., twelfth days after the assimilation). These simulations are forced by the new ECMWF analysis data that become available after the initial forecast run. New 10-day ocean forecasts are initialized at the end of each simulation. In this way, 10-day ocean forecasts are generated every day as the standard MFS output. Every fourteenth day, a new data assimilation step is performed and the MFS forecast stems directly from the initial conditions, without a simulation step. All other days, the 10-day ocean forecast launches from the assimilation period plus the simulations, where the simulation period increases by 1 day, everyday.

MFS Data Assimilation Step with MFS-Wind-BHM

In our reforecast experiments, we use 10 realizations from the posterior distribution for MFS-Wind-BHM to drive both the assimilation step ($t = 1, T$) and the 10-day forecast ($t = T, T + L$). The basin scale statistics of the 10 realizations are similar from realization to realization. Basin average RMS difference with respect to the ensemble mean are around $0.75\,ms^{-1}$; dropping to $0.55\,ms^{-1}$ when scatterometer data are available during the assimilation step.

Importantly, the ocean response to the MFS-Wind-BHM realizations during the assimilation step concentrates the largest RMS differences in mesoscale eddy structures. Figure 3a is a map of the initial condition (i.e. $t = 14d$) spread in SSH. Figure 3b is the initial condition spread in sea-surface temperature (SST). The mesoscale eddies are the dominant features of the initial condition spread.

This is an ideal response for the MFS data assimilation system, and it is a direct result of having forced the assimilation step with surface wind realizations from the MFS-Wind-BHM. Initial condition spread is a measure of the uncertainty in the assimilation procedure; uncertainty that stems from both observational and model errors. Figure 1 depicts uncertainty in the surface forcing that is sensitive to the region and meteorological scenario captured in the observations and analyses. Figure 3 implies that subtle changes in speed and direction of the surface forcing drive subtle differences in Ekman pumping and suction at the surface. These differences accumulate in mesoscale eddies that are the most uncertain features of the Mediterranean circulation. An initial condition perturbation method that concentrates uncertainty in the mesoscale eddies is therefore optimal.

MFS Forecast Step with MFS-Wind-BHM

A major goal of ensemble forecast methodologies is to create a set of closely-related, physically realizable initial states from which separate forecast integrations can begin. As the forecast integrations evolve, the ensemble members will spread (cluster) about uncertain (relatively certain) forecast states. The forecast ensemble response quantifies the uncertainty in each forecast state in terms of the spread (i.e. RMS differences).

The initial condition concentration of MFS ensemble forecast uncertainty remains in the mesoscale eddies for the duration of the MFS re-forecast experiments run to date. Figure 4 depicts the SSH spread at days 3 (Fig 4a) and 10 (Fig 4b) of the forecast period. The largest spread occurs in mesoscale patterns associated with the Algerian Current in the Western Basin, and with a boundary current off North Africa in the Eastern Basin. The Algerian Current instabilities are being analyzed in greater detail by Bonazzi et al. (2006).

Animations of 2 forecast realizations in the Gulf of Lyons were created for the presentation at ONR HQ. The animations demonstrate significantly different timings and quantities of deep water formation in the pre-conditioned region of the Gulf. This suggests a sub-basin to general circulation scale implication of the different forcings from the wind realizations drawn from the MFS-Wind-BHM posterior distribution. The sub-basin scale ocean response differences are due to the impact of different mesoscale evolutions, driven by subtle differences in the wind. Because only one ensemble member exhibits earlier and more abundant deep water formation, we can characterize this particular forecast as less likely, but nonetheless physically consistent within the bounds of uncertainty that propagate from the surface wind forcing.

The initial condition uncertainty distribution, and the retention of forecast uncertainties in the mesoscale eddies is a significant achievement of the MFS ensemble forecast system, given surface wind forcing realizations from MFS-Wind-BHM.

MFS-Error-BHM Plans for Year 2

In year 2 of the project we will investigate BHM methods for modelling the evolution of error covariance in MFS. In the MFS ROOI method, the full covariance matrix P is replaced by a background error covariance B. The forecast background error covariance matrix is decomposed

$$\mathbf{B}^f = \mathbf{D}^{f^{1/2}}\mathbf{C}^{-1}\mathbf{D}^{f^{1/2}} \tag{3}$$

where $\mathbf{D}^f$ is a diagonal matrix with diagonal elements quantifying the forecast error variances, and $\mathbf{C}$ is a spatial correlation matrix.

The evolution of the forecast error variances is *ad hoc*; we let $\mathbf{D}^f = \boldsymbol{N}\mathbf{D}^a$. MFS-Error-BHM will be developed to model $\boldsymbol{N}$.

References

Bonazzi, A. and R.F. Milliff, 2006: "Measuring predictability in MFS ensemble forecasts driven by BHM winds", in preparation.

Milliff, R.F., 2004: "Comparing Kinetic Energy vs. Wavenumber in Surface Wind Fields from ECMWF Analyses and the NASA QuikSCAT Scatterometer", Final Rpt. No. EVK3-CT-2002-00075 MFSTEP, 18 pgs.

Royle, J.A., L.M. Berliner, C.K. Wikle, R.F. Milliff, 1998: "A hierarchical spatial model for constructing wind fields from scatterometer data in the Labrador Sea", *Case Studies in Bayesian Statistics IV*, C. Gatsonis, R.E. Kass, B. Carlin, A. Cariquiry, A. Gelman, I. Verdinelli and M. West (eds.), Springer-Verlag, 367-381.

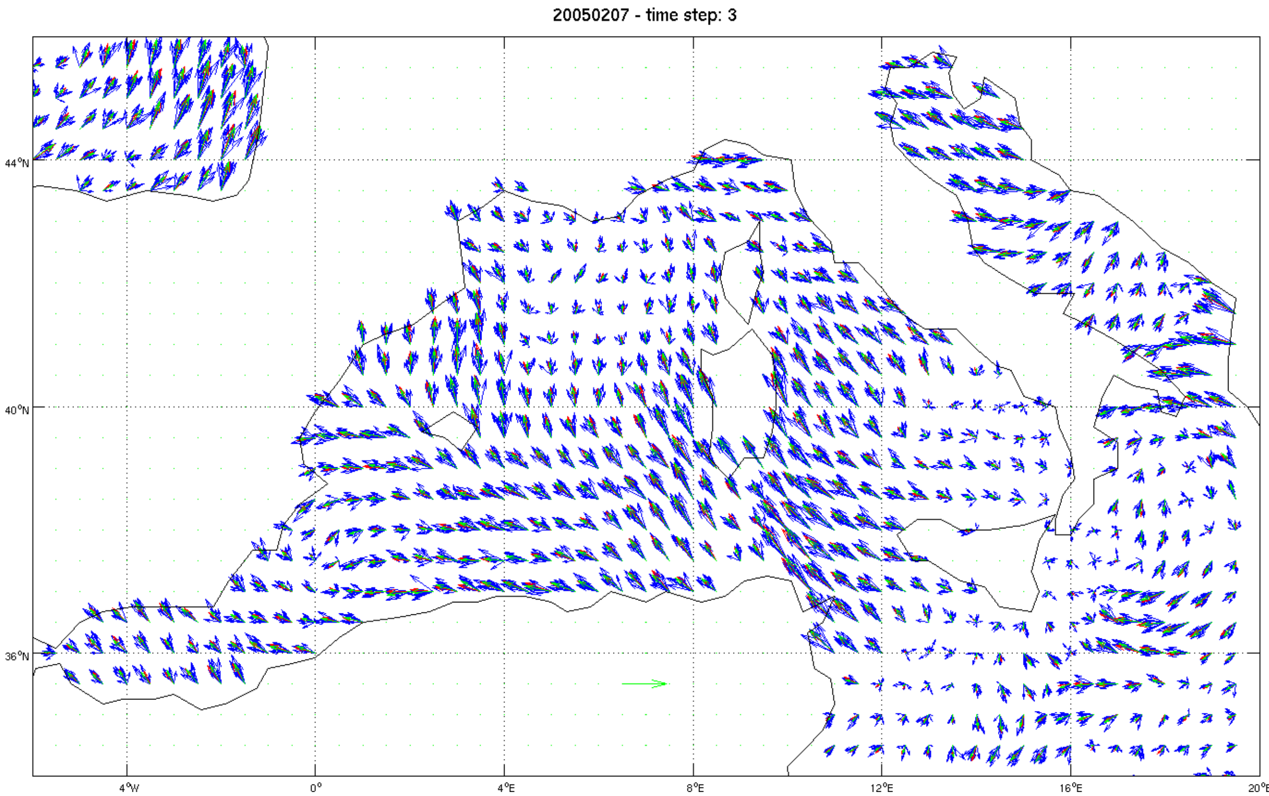

Figure 1: Snapshot of surface wind vector fields for the Western Mediterranean from MFS-Wind-BHM. Ten realizations from the posterior distribution are shown at each output grid point (blue vectors). The ensemble mean vector is green and a red vector denotes the ECMWF analysis vector for the same time. For clarity, wind vectors are plotted at every other grid point in the MFS-Wind-BHM prediction grid.

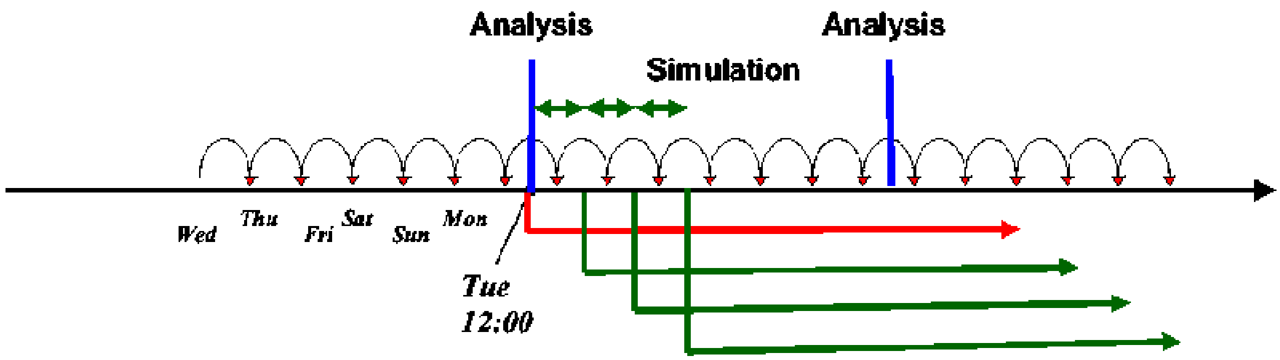

Figure 2: MFS data assimilation and forecast cycle. Full ocean data assimilation steps are run every 14 days (blue bars). A forecast is launched at the end of each assimilation step (red arrow). Forecasts are also launched on days between data assimilation steps, after a daily simulation run (green arrows), wherein only the atmospheric forcing is updated (i.e. ECMWF analyses replace atmospheric forecast fields).

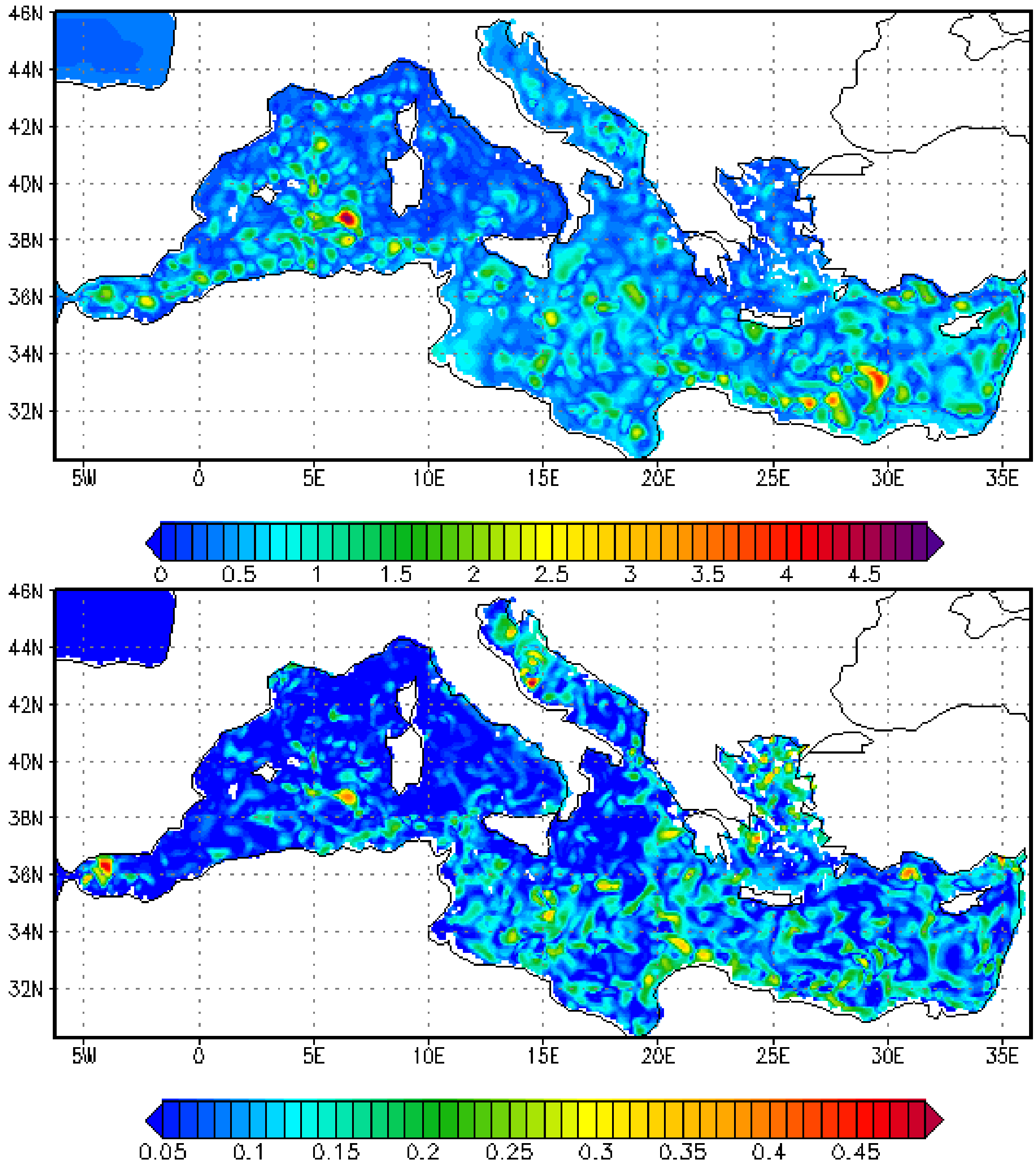

Figure 3: MFS ensemble initial condition spread (RMS differences) in sea surface height (SSH in cm; top panel) and sea surface temperature (SST in $^\circ C$; bottom panel). Initial condition differences are due to slight differences in wind forcing, for the data 14-day data assimilation step, from across the 10 realizations of the posterior distribution drawn from MFS-Wind-BHM. In both SSH and SST the differences accumulate in the mesoscale eddy field.

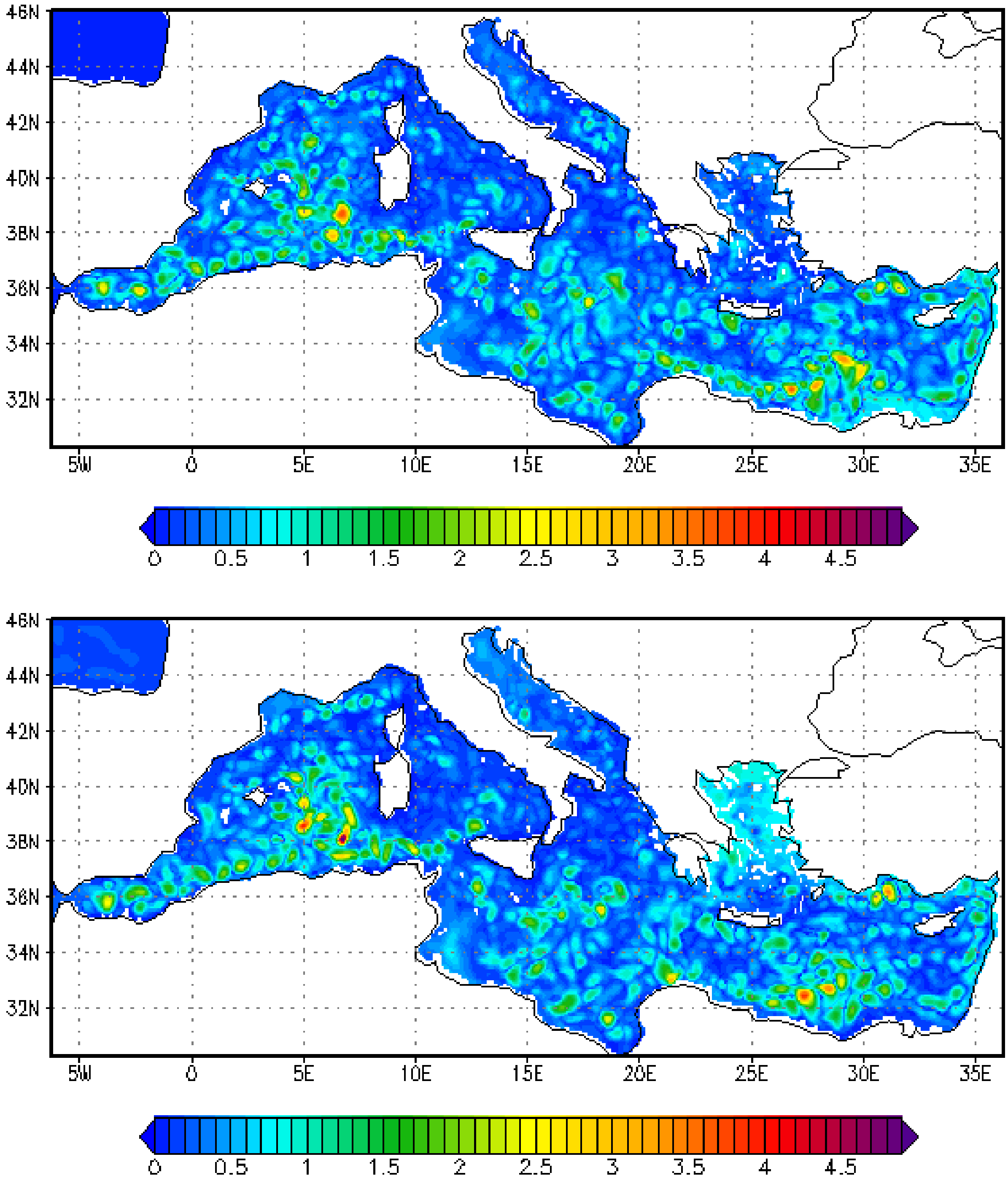

Figure 4: MFS forecast spread in SSH (cm) at forecast days 3 (top panel) and 10 (bottom panel). Ten MFS forecasts are forced by different realizations from the MFS-Wind-BHM posterior distribution. Forecasts are launched from corresponding initial conditions that were also generated by BHM winds in the data assimilation step. The initial condition spread that was concentrated in the oceanic mesoscale (see Fig. 3), persists at that scale throughout the 10-day forecasts.